Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Enzyme evolution: innovation is easy, optimization is complicated

Matilda S .Newton[1], Vickery L Arcus[2], Monica L Gerth[3] and Wayne M Patrick[3]

Enzymes have been evolving to catalyze new chemical reactions for billions of years, and will continue to do so for billions more. Here, we review examples in which evolutionary biochemists have used big data and high-throughput experimental tools to shed new light on the enormous functional diversity of extant enzymes, and the evolutionary processes that gave rise to it. We discuss the role that gene loss has played in enzyme evolution, as well as the more familiar processes of gene duplication and divergence. We also review insightful studies that relate not only catalytic activity, but also a host of other biophysical and cellular parameters, to organismal fitness. Finally, we provide an updated perspective on protein engineering, based on our new-found appreciation that most enzymes are sloppy and mediocre.

Addresses
[1] BioTechnology Institute, University of Minnesota, St. Paul, MN, United States
[2] School of Science, University of Waikato, Hamilton, New Zealand
[3] Department of Biochemistry, University of Otago, Dunedin, New Zealand

Corresponding author: Patrick, Wayne M (wayne.patrick@otago.ac.nz)

Current Opinion in Structural Biology 2018, **48**:110–116

This review comes from a themed issue on **Proteins: an evolutionary perspective**

Edited by **Birte Höcker** and **Jakob R Winther**

For a complete overview see the Issue and the Editorial

Available online 5th December 2017

https://doi.org/10.1016/j.sbi.2017.11.007

0959-440X/© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

New enzymes have been evolving on Earth for at least four billion years, and will continue to do so for another two billion or so — at which point the expanding sun will sterilize our planet [1]. The goal of this article is to review recent studies that shed new light on enzyme evolution, with a focus on work published since 2015.

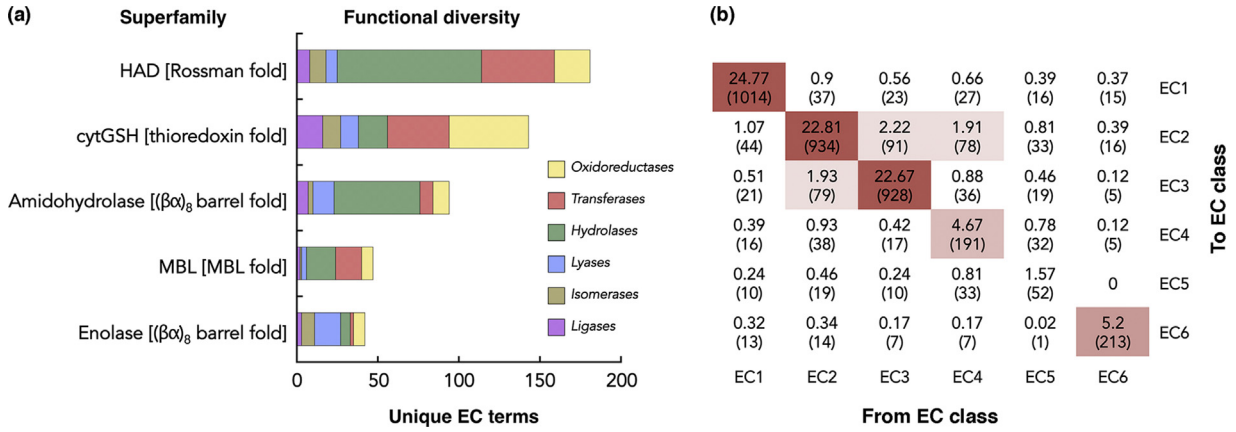
Innovation is easy

A general model for the evolution of enzymes with new functions was articulated by Yčas and Jensen, independently, in the mid-1970s [2,3]. Each proposed that ancestral enzymes were multifunctional generalists, with the ability to catalyze broad classes of reactions on a range of substrates. From this low-fidelity starting point, gene duplication and divergence would have given rise to more specialized enzymes with higher activities towards their preferred substrates.

The Yčas–Jensen model had two important implications. First, divergent evolution of new enzymes was most likely to be enabled, and constrained, by catalytic chemistry. Gerlt and Babbitt were among the first to emphasize the importance of 'chemistry driven' evolution from multifunctional ancestors, giving rise to superfamilies of homologous enzymes. As originally defined, the members of these superfamilies share the same fold and either catalyze the same reaction with different substrate specificities, or catalyze different overall reactions that share a common mechanistic feature such as a partial reaction, an intermediate or a transition state [4]. The second implication of the Yčas–Jensen model was that the promiscuous (secondary, non-physiological) activities of existing enzymes remain important starting points for the evolution of new functions, because today's enzymes are tomorrow's ancestors. It is now well accepted that most — and probably all — extant enzymes are, in fact, promiscuous [5,6].

Recent large-scale studies, both computational and experimental, have opened our eyes to the enormous functional diversity among existing enzyme superfamilies, the vastness of 'promiscuity space,' and therefore the seemingly limitless potential for future evolutionary innovation. Baier *et al.* surveyed the functional diversity, as represented by Enzyme Commission (EC) numbers, in five common superfamilies [7•]. Each superfamily contained enzymes from all six of the EC classes (Figure 1a). Furnham *et al.* went further and used a phylogenetic approach [8] to reconstruct the evolutionary histories of 379 superfamilies from the Class, Architecture, Topology, Homology (CATH) database, and to ask how often a change in EC number was observed over the course of their evolution [9•]. While 81% of the functional changes were within an EC class, every possible change between EC classes was also observed (Figure 1b), with the exception of a change from a ligase (EC class 6) to an isomerase (EC class 5). These bioinformatics studies emphasize that there is little, if anything, that constrains particular catalytic chemistries to particular folds.

Figure 1



Functional diversity in enzyme superfamilies. **(a)** Five abundant superfamilies, each with 13 000–91 000 representatives in the sequence databases, are HAD (haloalkanoate dehalogenase), cytGST (cytosolic glutathione transferase), amidohydrolase, MBL (metallo-β-lactamase), and enolase. Each of these superfamilies contains homologous enzymes that fall into all six EC classes. Adapted with permission from [7•]. Copyright (2016) American Chemical Society. **(b)** An EC exchange matrix, counting the changes from one EC number to another during the evolution of 379 different superfamilies. Counts are expressed as a percentage of the total number observed, with the raw numbers of exchanges in parentheses. Colouring is on a red intensity scale. Reproduced with modifications from [9•].

Four high-throughput experimental studies (reviewed in detail elsewhere [7•]) have reached a similar conclusion. Dozens of enzymes from within the cytosolic glutathione transferase [10], β-keto acid cleavage enzyme [11], metallo-β-lactamase [12], and haloalkanoate dehalogenase [13••] superfamilies were each tested for activity towards a range of different substrates. In each case, many enzymes were found to have multiple functions *in vitro*. In the most comprehensive study, 217 members of the haloalkanoate dehalogenase superfamily were expressed, purified, and screened for phosphatase or phosphonatase activity towards 167 substrates (most of which were naturally occurring metabolites). The authors discovered breathtakingly broad substrate specificities. A median of 15.5 substrates were recognized by each enzyme, 50 of the enzymes could utilize 40 or more substrates, and remarkably, one enzyme could utilize 143 [13••].

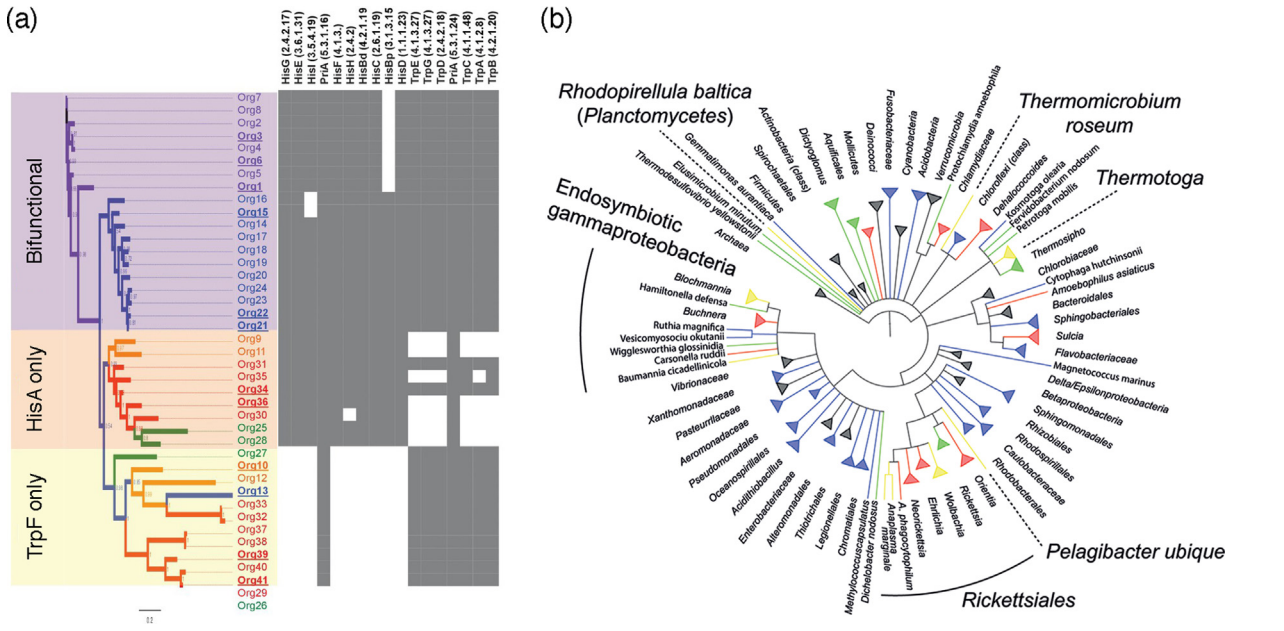
Together, these computational and experimental studies highlight the genuine risk in using homology to assign physiological function(s) to uncharacterized proteins in databases [14]. A further caveat with *in vitro* experiments is that it can be difficult to elucidate which activities are physiological (being maintained by selection) and which are promiscuous. Even in cases where an enzyme appears to have a clear-cut physiological role, it is theoretically possible that one or more of its weak side activities may be contributing to the fitness of the organism — either by contributing to the metabolite pool, or by inducing a regulatory effect. Regardless, the old idea of 'one enzyme, one substrate' is now shown to be quaint and outdated. The leading evolutionary biochemist, Prof Shelley Copley, has made the entirely reasonable estimate that an average enzyme may have 10 promiscuous activities [6]. Thus, even the simplest bacteria are likely to harbour 10 000–20 000 promiscuous activities, any one of which may be the starting point for the evolution of a new enzyme. Not only that, but two upcoming studies have retraced the evolution of enzymes from their non-catalytic ancestors, via a small number of key mutations in each case [15,16]. When non-enzymatic scaffolds are also considered, there is certainly no shortage of possibilities for future evolutionary innovation!

Gene loss drives functional innovation, too

The Yčas–Jensen model and its descendants (e.g. [17,18]) are centred on the importance of gene duplication and divergence as the evolutionary route to new enzymes. However, genome reduction is also a pervasive force in evolution. Every lineage, apart from that tiny fraction leading to extant animals and plants, appears to undergo rapid bursts of genomic complexification, followed by much longer periods in which genetic material is slowly lost [19]. Two recent studies have combined phylogenomics and biochemistry to examine how gene loss can shape enzyme evolution.

Juárez-Vázquez *et al.* [20••] continued their groundbreaking research into the evolution of PriA, a bifunctional isomerase that catalyzes the HisA and TrpF reactions (in histidine and tryptophan biosynthesis, respectively) in some bacteria [21–23]. Extensive phylogenomic analysis and the construction of genome-scale metabolic models for 33 bacteria led to the identification of PriA homologues that were predicted to fulfil different roles (bifunctional, HisA-only, or TrpF-only), depending on the pattern of gene loss in the host organism (Figure 2a).

Figure 2



Phylogenomics to inform studies of enzyme evolution. **(a)** A phylogenetic tree of PriA sequences from the bacterial family Actinomycetaceae. When compared with the occurrence of enzymes for histidine (His) and tryptophan (Trp) biosynthesis in the same species, it becomes apparent that PriA has co-evolved with the presence (grey) or absence (white) of these enzymes. Figure reproduced with modification from [20••]. **(b)** A survey of bacterial genomes for genes encoding CBL and ALR. Taxa in green have both enzymes. Taxa in blue lack CBL, but have ALR. Taxa in grey have ALR, but have mixed occurrence of CBL. Taxa in red lack both enzymes. Taxa in yellow and labelled in larger type have CBL but lack ALR. These latter taxa were candidates for cases in which CBL had evolved to be a bifunctional CBL/ALR. Figure reproduced from [24•].

Kinetics and structural biology were used to confirm and explain these predictions. The bifunctional PriA ancestor had evolved to become a monofunctional HisA in species that had lost their *trp* genes, and a monofunctional TrpF in species that had lost their *his* genes. This study is important because it explores a scenario in which evolution of monofunctional enzymes was driven by the relaxation of purifying selection (i.e. no further need to retain the second function), rather than the more well-studied situation in which there is positive selection *for* a function. This version of adaptation — which has undoubtedly been prevalent during evolution by genome reduction — yielded specialist enzymes with poor kinetic parameters, especially with respect to turnover ($k_{cat} < 0.2\ s^{-1}$).

On the other hand, we examined a case in which gene loss has led to the evolution of multifunctional enzymes from specialist ancestors [24•]. We previously showed that the *Escherichia coli* cystathionine β-lyase (CBL) has promiscuous alanine racemase (ALR) activity, even though the two enzymes share no sequence or structural similarity [25]. We used this insight to search >1000 bacterial genomes for those with a gene encoding CBL but no ALR. Our hypothesis was that CBL may have evolved to become a bifunctional CBL/ALR in these species. The hallmark of the taxa that met our criterion (Figure 2b), is that they all have reduced genomes. We characterized the CBLs from *Pelagibacter ubique*, the *Wolbachia* endosymbiont of *Drosophila melanogaster*, and *Thermotoga maritima*, all of which had the two predicted activities *in vitro*. The *Wolbachia* and *T. maritima* enzymes also had glutamate racemase activity. However, we also showed that CBL activity is no longer physiologically relevant in these bacteria. Instead, the gain of a multifunctional CBL by horizontal transfer has enabled the subsequent loss of the genes for alanine racemase and glutamate racemase. Like Juárez-Vázquez *et al.*, we measured particularly low activities (k_{cat}/K_M values of 6.3–690 $s^{-1}\ M^{-1}$) for the physiological reactions catalyzed by these CBL enzymes, which have been caught in the act of evolving into broad-specificity amino acid racemases.

These examples emphasize that many microorganisms continue to use primordial-like enzymes with broad specificities and/or weak activities in key metabolic roles. Indeed, bacteria such as *Wolbachia* (thought to infect most insects) and *P. ubique* (a marine bacterium) are among the most abundant organisms on the planet [26]. It is possible that the primordial-like enzymes that were studied [20••,24•] might be short-lived evolutionary intermediates, and that their kinetic parameters will improve in future. However, preliminary analyses with the TimeTree server (www.timetree.org) suggest that tens of millions, to hundreds of millions, of years have passed since the gene loss events that drove their evolution. Further

phylogenetics analyses, and broader searches for similar enzymes in other species, will help to resolve this question. For now, it appears safe to conclude that the evolutionary success of a species is not correlated with the kinetic parameters of its central metabolic enzymes.

How does selection for organismal fitness affect evolving enzymes?

Selection acts at the level of the organism (and the population). Increasingly, it is being recognized that connecting enzyme evolution to organismal fitness (i.e. the ability to leave viable offspring) is both: (a) required if we are to gain a more accurate and holistic understanding of evolutionary processes; and (b) an outstanding challenge [27–30].

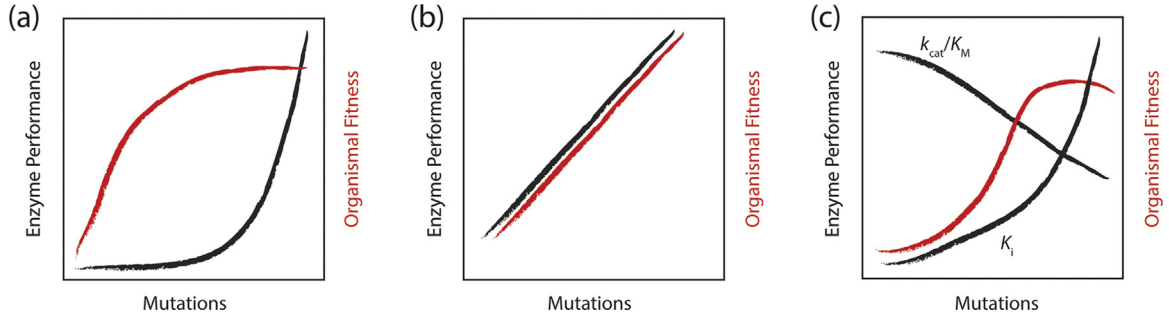
A number of insightful recent papers have studied the evolution of enzyme activity within the context of organismal fitness. This flurry of recent activity is making use of new and high-throughput technologies, although as with many aspects of evolutionary biology, the intellectual framework was built decades ago, by pioneers such as Kacser and Dean [31–35].

The first important point when relating enzyme activity to organismal fitness is that small functional changes have the greatest fitness effects when they occur in a poor enzyme, as any change in the flux through a rate-limiting metabolic step is reflected in cell growth. On the other hand, even substantial alterations to enzymes with high activity (near, at or above the threshold required by the cell) have negligible effects at the pathway or organism level [31,33]. This is exactly what we observed when we characterized the structures and functions of HisA variants with newly evolved TrpF activity [36•]. These new TrpF enzymes were isolated in a previous real-time evolution experiment with *Salmonella enterica* [37]. Small improvements to enzyme performance (defined as $k_{cat}/K_M \times$ relative protein abundance) early in the evolutionary trajectory led to marked improvements in cell growth rate; whereas a later mutation that elicited the greatest increase in enzyme performance had little effect on fitness (Figure 3a).

Klesmith *et al.* observed a similar phenomenon when they optimized a synthetic metabolic pathway for levoglucosan utilization in *E. coli* [38•]. A first-round mutant of levoglucosan kinase, LGK.1, showed only a marginal, 7% increase in catalytic efficiency over wildtype (and a 5°C increase in T_m), but conferred a 28-fold increase in growth rate when levoglucosan was the sole carbon source. The final mutant, LGK.9, had double the k_{cat}/K_M, but only conferred an extra 1.3-fold improvement in growth rate over LGK.1. This comprehensive study shows that understanding the relationship between enzyme activity, metabolic flux and organismal fitness will be important for realizing the potential of synthetic biology.

The second important point when considering activity and fitness is that kinetic parameters such as k_{cat} or k_{cat}/K_M are not sufficient as the sole definition of an enzyme's contribution to fitness. For a start, kinetics data are rarely collected under conditions that mimic those encountered physiologically — although ideally, they would be [39]. Kinetics data also need to be scaled by the level of soluble, functional enzyme [29,36•], and cofactor availability may be an additional consideration [35]. An illustrative example of these factors was recently provided by Meini *et al.* [40•]. They studied the mutational trajectory

Figure 3



Different relationships between the evolution of enzyme performance (left axes; black) and organismal fitness (right axes; red). **(a)** Enhancing enzyme performance improves fitness, to a point. The final TrpF mutation selected in a real-time evolution experiment using *S. enterica* increased enzyme performance seven-fold, but had a negligible effect on growth rate [36•]. **(b)** Enzyme performance directly affects cell fitness. A metallo-β-lactamase was evolved to confer cephalexin resistance. Fitness (i.e. MIC for cephalexin) was directly proportional to enzyme performance, but only when enzyme variants were tested under physiological conditions [40•]. **(c)** Enzyme activity and inhibition trade off as fitness increases. *E. coli* was evolved for resistance towards trimethoprim, which acts by competitively inhibiting DHFR. Mutations that limited antibiotic binding (increased K_i) also reduced enzyme stability and activity (k_{cat}/K_M shown). Initially, the mutations had beneficial effects on cell fitness (IC_{50}) by alleviating the effect of the antibiotic. Interestingly, fitness diminished at the end of the trajectory, once enzyme activity was sufficiently weakened to become growth-limiting [41••].

leading to an artificially evolved metallo-β-lactamase, which conferred high levels of resistance to cephalexin. Ultimately, they discovered that enzyme performance was highly correlated with the minimum inhibitory concentration (MIC) for the antibiotic; that is, with organismal fitness (Figure 3b). However, this conclusion was only reached after metallo-β-lactamase variants were assayed in the physiologically relevant context, which was periplasmic extract without the Zn^{2+} that is usually added as a supplement to the assay buffer. There was no correlation between the catalytic efficiencies of purified enzyme variants (determined in the standard buffer with excess Zn^{2+}) and their effect on MIC. Instead, selection for cephalexin resistance had acted to hone a combination of catalytic efficiency and Zn^{2+} binding affinity in the evolving enzyme.

Going even further, Rodrigues *et al.* explored the evolution of dihydrofolate reductase (DHFR) to confer trimethoprim resistance on *E. coli* [41••]. In this scenario, the enzyme must maintain its essential catalytic activity while accumulating mutations that prevent its inhibition by the antibiotic. For each possible combination of three mutations, the authors characterized biophysical parameters including catalytic efficiency, the inhibition constant (K_i), thermostability and the propensity to form molten globules, as well as intracellular protein abundance and the cellular resistance phenotype (IC_{50}). Progressive mutations not only led to a trade-off between inhibitor binding and catalytic efficiency (Figure 3c), but also modulated the concentration of functional DHFR in the cell through stability and protein quality control mechanisms. Organism-level selection had acted upon the global effects of these combined factors. Impressively, the authors were also able to derive a robust predictive function for organismal fitness, based solely upon molecular properties of the enzyme variants (k_{cat}/K_M, K_i and bis-ANS fluorescence) that were measured *in vitro*.

Concluding remarks and a new perspective on enzyme engineering

Six years ago, a landmark survey of all kinetic parameters in the BRENDA database revealed that the 'average enzyme' was moderately efficient, with a k_{cat} of $\sim 10\ s^{-1}$ and a k_{cat}/K_M of $\sim 10^5\ s^{-1}\ M^{-1}$ [42]. Here, we have highlighted some of the work that explains and updates this overarching view of enzymology. The evolution of new functions, and the re-emergence of old functions on new folds, has been relentless, over billions of years. Dynamic genomes — forever gaining and losing genetic material — lead to pressures and constraints on the evolution of enzymes that are gained or retained. Critically, the contribution of an enzyme to metabolic flux and organismal fitness is not limited to its k_{cat} or k_{cat}/K_M, but instead its evolutionary trajectory is determined by a combination of biochemical, biophysical and regulatory factors.

Thus, it is not surprising that most enzymes are far from perfect catalysts — evolution is not the pursuit of perfect enzymes. Indeed, we have highlighted studies that show the opposite [20••,24•,36•]. Abundant but understudied microorganisms make use of enzymes that are orders of magnitude worse than the average enzyme which is currently in BRENDA. 'Real world' enzymes are sloppy and mediocre. However, protein engineers often bemoan their perceived inability to design or evolve catalysts that rival those found in nature [43,44]. One heart-warming aspect of our discussion is that they need not be so hard on themselves! Moreover, engineering strategies that build on our improving knowledge of enzyme evolution are beginning to emerge.

The idea that innovation is easy when starting from a promiscuous enzyme scaffold has been successfully utilized by many groups for the directed evolution of new functions [45]. Khanal *et al.* refined this idea, proposing that directed evolution experiments should start with more than one scaffold [46•]. Their rationale was based on the discovery that levels of a promiscuous activity vary amongst orthologues, and that the effects of mutations to improve this activity do not necessarily correlate with its initial level. A related strategy, termed scaffold sampling, has also been explored [47]. In this approach, beneficial mutations are identified in one enzyme, and then transposed onto a series of homologous scaffolds.

Finally, we note that 1003 phylogenetically diverse reference genomes have recently been released by the Genomic Encyclopaedia of Bacteria and Archaea (GEBA) project [48]. This exciting resource offers almost 500 000 entirely new protein sequence clusters, waiting to be characterized and used by protein engineers. It will also provide researchers with interests in genomic enzymology [14], esoteric enzymology [49], phyloenzymology [24•] and evolutionary biochemistry [50] ample opportunity to gain a newly realistic view of enzyme evolution and metabolic biochemistry throughout the biosphere. There has never been a better time to study enzyme evolution!

Acknowledgements

This work was supported by the Marsden Fund (VLA/MLG/WMP) and by Smart Ideas grants from the New Zealand Ministry of Business, Innovation and Employment (VLA/MLG/WMP). WMP is the grateful recipient of a Rutherford Discovery Fellowship.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Brewster JL, Finn TJ, Ramirez MA, Patrick WM: **Whither life? Conjectures on the future evolution of biochemistry**. *Biol Lett* 2016, **12**:20160269.

2. Yčas M: **On earlier states of the biochemical system**. *J Theor Biol* 1974, **44**:145-160.

3. Jensen RA: **Enzyme recruitment in evolution of new function**. *Annu Rev Microbiol* 1976, **30**:409-425.

4. Gerlt JA, Babbitt PC: **Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies**. *Annu Rev Biochem* 2001, **70**:209-246.

5. Khersonsky O, Tawfik DS: **Enzyme promiscuity: a mechanistic and evolutionary perspective**. *Annu Rev Biochem* 2010, **79**:471-505.

6. Copley SD: **Shining a light on enzyme promiscuity**. *Curr Opin Struct Biol* 2017, **47**:167-175.

7. • Baier F, Copp JN, Tokuriki N: **Evolution of enzyme superfamilies: comprehensive exploration of sequence-function relationships**. *Biochemistry* 2016, **55**:6375-6388.
A survey of the functional diversity in several well-known superfamiles, and an insightful review of large-scale experiments to investigate superfamily evolution.

8. Furnham N, Sillitoe I, Holliday GL, Cuff AL, Rahman SA, Laskowski RA, Orengo CA, Thornton JM: **FunTree: a resource for exploring the functional evolution of structurally defined enzyme superfamilies**. *Nucleic Acids Res* 2012, **40**:D776-D782.

9. • Furnham N, Dawson NL, Rahman SA, Thornton JM, Orengo CA: **Large-scale analysis exploring evolution of catalytic machineries and mechanisms in enzyme superfamilies**. *J Mol Biol* 2016, **428**:253-267.
The high frequency of functional changes during enzyme evolution is uncovered by reconstructing the evolutionary histories of 379 superfamilies (with respect to the functions of their members).

10. Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, Jagessar K, Kim J, Patskovsky Y, Seidel RD *et al.*: **Large-scale determination of sequence, structure, and function relationships in cytosolic glutathione transferases across the biosphere**. *PLoS Biol* 2014, **12**:e1001843.

11. Bastard K, Smith AA, Vergne-Vaxelaire C, Perret A, Zaparucha A, De Melo-Minardi R, Mariage A, Boutard M, Debard A, Lechaplais C *et al.*: **Revealing the hidden functional diversity of an enzyme family**. *Nat Chem Biol* 2014, **10**:42-49.

12. Baier F, Tokuriki N: **Connectivity between catalytic landscapes of the metallo-β-lactamase superfamily**. *J Mol Biol* 2014, **426**:2442-2456.

13. •• Huang H, Pandya C, Liu C, Al-Obaidi NF, Wang M, Zheng L, Toews Keating S, Aono M, Love JD, Evans B *et al.*: **Panoramic view of a superfamily of phosphatases through substrate profiling**. *Proc Natl Acad Sci U S A* 2015, **112**:E1974-E1983.
Comprensive functional profiling of >200 enzymes with >150 substrates reveals the vast extent of multifunctionality in the haloalkanoate dehalogenase superfamily.

14. Gerlt JA: **Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions**. *Biochemistry* 2017, **56**:4293-4308.

15. Clifton BE, Kaczmarski JA, Carr PD, Gerth ML, Tokuriki N, Jackson CJ: **Evolution of an enzyme from a solute-binding protein**. *bioRxiv* 2017 http://dx.doi.org/10.1101/157495.

16. Kaltenbach M, Burke JR, Dindo M, Pabis A, Munsberg FS, Rabin A, Kamerlin SCL, Noel JP, Tawfik DS: **Evolution of chalcone isomerase from a non-catalytic ancestor**. *bioRxiv* 2017 http://dx.doi.org/10.1101/174128.

17. Bergthorsson U, Andersson DI, Roth JR: **Ohno's dilemma: evolution of new genes under continuous selection**. *Proc Natl Acad Sci U S A* 2007, **104**:17004-17009.

18. Sikosek T, Chan HS, Bornberg-Bauer E: **Escape from Adaptive Conflict follows from weak functional trade-offs and mutational robustness**. *Proc Natl Acad Sci U S A* 2012, **109**:14888-14893.

19. Wolf YI, Koonin EV: **Genome reduction as the dominant mode of evolution**. *Bioessays* 2013, **35**:829-837.

20. •• Juárez-Vázquez AL, Edirisinghe JN, Verduzco-Castro EA, Michalska K, Wu C, Noda-García L, Babnigg G, Endres M, Medina-Ruíz S, Santoyo-Flores J *et al.*: **Evolution of substrate specificity in a retained enzyme driven by gene loss**. *eLife* 2017, **6**:e22679.
A *tour de force*: phylogenomics, metabolic modelling, phylogenetics, genome sequencing, microbiology, kinetics and X-ray crystallography reveal how gene loss can drive the evolution of retained enzymes.

21. Barona-Gómez F, Hodgson DA: **Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis**. *EMBO Rep* 2003, **4**:296-300.

22. Noda-García L, Camacho-Zarco AR, Medina-Ruíz S, Gaytán P, Carrillo-Tripp M, Fülöp V, Barona-Gómez F: **Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer**. *Mol Biol Evol* 2013, **30**:2024-2034.

23. Verduzco-Castro EA, Michalska K, Endres M, Juárez-Vazquez AL, Noda-García L, Chang C, Henry CS, Babnigg G, Joachimiak A, Barona-Gómez F: **Co-occurrence of analogous enzymes determines evolution of a novel $(\beta\alpha)_8$-isomerase sub-family after non-conserved mutations in flexible loop**. *Biochem J* 2016, **473**:1141-1152.

24. • Ferla MP, Brewster JL, Hall KR, Evans GB, Patrick WM: **Primordial-like enzymes from bacteria with reduced genomes**. *Mol Microbiol* 2017, **105**:508-524.
Examining 1023 bacterial genomes for the presence and absence of genes associated with methionine and peptidoglycan biosynthesis led to the identification of multifunctional enzymes, which resemble those envisaged by Yčas and Jensen.

25. Soo VWC, Yosaatmadja Y, Squire CJ, Patrick WM: **Mechanistic insights from the reciprocal promiscuity and directed evolution of two PLP-dependent enzymes**. *J Biol Chem* 2016, **291**:19873-19887.

26. Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ: **SAR11 clade dominates ocean surface bacterioplankton communities**. *Nature* 2002, **420**:806-810.

27. Kaltenbach M, Tokuriki N: **Dynamics and constraints of enzyme evolution**. *J Exp Zool B Mol Dev Evol* 2014, **322**:468-487.

28. Newton MS, Arcus VL, Patrick WM: **Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes**. *J R Soc Interface* 2015, **12**:20150036.

29. Boucher JI, Bolon DN, Tawfik DS: **Quantifying and understanding the fitness effects of protein mutations: laboratory versus nature**. *Protein Sci* 2016, **25**:1219-1226.

30. Bershtein S, Serohijos AW, Shakhnovich EI: **Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations**. *Curr Opin Struct Biol* 2017, **42**:31-40.

31. Kacser H, Burns JA: **The molecular basis of dominance**. *Genetics* 1981, **97**:639-666.

32. Kacser H, Beeby R: **Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection**. *J Mol Evol* 1984, **20**:38-51.

33. Hartl DL, Dykhuizen DE, Dean AM: **Limits of adaptation: the evolution of selective neutrality**. *Genetics* 1985, **111**:655-674.

34. Dykhuizen DE, Dean AM, Hartl DL: **Metabolic flux and fitness**. *Genetics* 1987, **115**:25-31.

35. Lunzer M, Miller SP, Felsheim R, Dean AM: **The biochemical architecture of an ancient adaptive landscape**. *Science* 2005, **310**:499-501.

36. • Newton MS, Guo X, Söderholm A, Näsvall J, Lundström P, Andersson DI, Selmer M, Patrick WM: **Structural and functional innovations in the real-time evolution of new $(\beta\alpha)_8$ barrel enzymes**. *Proc Natl Acad Sci U S A* 2017, **114**:4727-4732.
A detailed examination of how selection for organismal fitness honed the structure, function, dynamics and expression levels of HisA as it evolved to possess TrpF activity.

37. Näsvall J, Sun L, Roth JR, Andersson DI: **Real-time evolution of new genes by innovation, amplification, and divergence**. *Science* 2012, **338**:384-387.

38. Klesmith JR, Bacik JP, Michalczyk R, Whitehead TA:
• **Comprehensive sequence-flux mapping of a levoglucosan utilization pathway in *E. coli***. *ACS Synth Biol* 2015, **4**:1235-1243.
A smart new pathway engineering approach for synthetic biologists, which incorporates measurements of activity, flux and fitness.

39. van Eunen K, Bakker BM: **The importance and challenges of *in vivo*-like enzyme kinetics**. *Perspect Sci* 2014, **1**:126-130.

40. Meini MR, Tomatis PE, Weinreich DM, Vila AJ: **Quantitative**
• **description of a protein fitness landscape based on molecular features**. *Mol Biol Evol* 2015, **32**:1774-1787.
A powerful demonstration of why it is important to measure enzyme performance under physiologically relevant conditions, in order to understand evolutionary trajectories.

41. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL,
•• Shakhnovich EI: **Biophysical principles predict fitness landscapes of drug resistance**. *Proc Natl Acad Sci U S A* 2016, **113**:E1470-E1478.
Deconstruction of an evolutionary trajectory into enough component parts — both molecular and cellular — that organismal fitness could be predicted based on *in vitro* measurements of enzyme properties.

42. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R: **The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters**. *Biochemistry* 2011, **50**:4402-4410.

43. Baker D: **An exciting but challenging road ahead for computational enzyme design**. *Protein Sci* 2010, **19**:1817-1819.

44. Kiss G, Celebi-Ölçüm N, Moretti R, Baker D, Houk KN: **Computational enzyme design**. *Angew Chem Int Ed Engl* 2013, **52**:5700-5725.

45. Renata H, Wang ZJ, Arnold FH: **Expanding the enzyme universe: accessing non-natural reactions by mechanism-guided directed evolution**. *Angew Chem Int Ed Engl* 2015, **54**:3351-3367.

46. Khanal A, Yu McLoughlin S, Kershner JP, Copley SD: **Differential**
• **effects of a mutation on the normal and promiscuous activities of orthologs: implications for natural and directed evolution**. *Mol Biol Evol* 2015, **32**:100-108.
An elegant experiment (obvious in hindsight!) that promises to change the way directed evolution is done.

47. Dunn MR, Otto C, Fenton KE, Chaput JC: **Improving polymerase activity with unnatural substrates by sampling mutations in homologous protein architectures**. *ACS Chem Biol* 2016, **11**:1210-1219.

48. Mukherjee S, Seshadri R, Varghese NJ, Eloe-Fadrosh EA, Meier-Kolthoff JP, Göker M, Coates RC, Hadjithomas M, Pavlopoulos GA, Paez-Espino D *et al.*: **1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life**. *Nat Biotechnol* 2017, **35**:676-683.

49. Tawfik DS, van der Donk WA: **Biocatalysis and biotransformation: esoteric, niche enzymology**. *Curr Opin Chem Biol* 2016, **31**:v-vii.

50. Copley SD: **An evolutionary biochemist's perspective on promiscuity**. *Trends Biochem Sci* 2015, **40**:72-78.